Exploring the Universe with Artificial Intelligence (UniversAI)
Proceedings IAU Symposium No. 397, 2026
I. Liodakis & V. Efthymiou, eds.
doi:10.1017/S1743921326106255

Investigation of Galactic Populations and Merger Process Using Machine Learning Methods

Furkan Akbaba

Institute of Graduate Studies Science, Istanbul University, Istanbul, Turkey
email: furkan.akbaba@ogr.iu.edu.tr

Abstract. We combine APOGEE DR17 spectroscopy with Gaia DR3 astrometry to investigate the chemo-dynamical structure of the Milky Way using unsupervised machine learning. After applying strict quality cuts, we selected a high-quality sample of 135,080 giant stars and derived their orbital parameters in the McMillan potential. To mitigate disk dominance, a cut of $e_p > 0.25$ was applied. Dimensionality reduction with PCA and clustering with K-means were tested across three input spaces: chemical abundances, kinematic parameters, and their combination. The most robust separation of stellar populations is obtained when using chemical abundances alone. The resulting clusters correspond to the thin disk, thick disk, halo, and an accreted component, as revealed in chemical-abundance and Lindbland diagrams. Our results highlight the effectiveness of abundance-driven clustering in disentangling Galactic stellar populations and tracing ex-situ stars, providing a data-driven framework for future studies of Galactic archeology.

Keywords. Milky Way Galaxy(1054), Milky Way formation(1053), Milky Way evolution(1052), Galaxy chemical evolution(580), Dwarf galaxies(416)

1. Introduction

The formation and present-day structure of the Milky Way are the cumulative outcome of both internal evolutionary processes and external accretion events (Bland-Hawthorn and Gerhard (2016)). Over cosmic time, dwarf galaxies and globular clusters have merged with the Galaxy, leaving behind stellar populations with distinct chemical and dynamical signatures. Several such accreted structures have been reported in the literature, including the Helmi streams (Helmi et al. 1999), Sequoia (Barbá et al. 2019; Matsuno et al. 2019; Myeong et al. 2019), Thamnos (Koppelman et al. 2019), LMS (Yuan et al. 2020), Pontus (Malhan et al. 2022), Sagittarius (Ibata et al. 1994), Arjuna and I'itoi (Naidu et al. 2020), and the Gaia-Sausage/Enceladus event (Helmi et al. 2018; Belokurov et al. 2018; Haywood et al. 2018; Mackereth et al. 2019). Stars that were accreted from outside the Galaxy are referred to as ex-situ, while those that formed within the Galaxy are known as in-situ stars. Different approaches have been developed to distinguish these populations in chemical and dynamical planes. For example, (Horta et al. 2023) demonstrated that the [Mg/Mn]-[Al/Fe] plane can effectively separate high-α, low-α disk stars (in-situ) from ex-situ populations. Similarly, (Plevne et al. 2020) showed a classification scheme in the [α/Fe]-[Fe/H] plane. (Belokurov and Kravtsov 2024) further proposed that boundaries drawn in the energy-angular momentum (E-L_z) diagram, using globular clusters, can separate in-situ from ex-situ populations. However, orbital and kinematical parameters are subject to dynamical heating, radial migration, and secular processes, and therefore may not reliably reveal the origin of stars on their

© The Author(s), 2026. Published by Cambridge University Press on behalf of The International Astronomical Union. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

https://doi.org/10.1017/S1743921326106255 Published online by Cambridge University Press

own (Plevne and Akbaba 2025). In contrast, chemical abundances are more resilient tracers, as they directly preserve the nucleosynthetic history of the environment where the stars formed (Şentürk et al. 2024).

Large spectroscopic surveys such as APOGEE DR17 (Abdurro'uf et al. 2022), complemented by precise astrometric data from Gaia (Gaia Collaboration et al. 2023), provide an unprecedented opportunity to investigate these questions on statistically robust scales. However, the high-dimensional complexity of these datasets often leads to overlaps in chemical and kinematic spaces. In this context, unsupervised machine learning techniques can be used to detect substructures and uncover hidden stellar populations. Dimensionality reduction methods such as Principal Component Analysis (PCA) (Jolliffe 1990), combined with clustering algorithms particularly K-Means (Jin and Han 2011), allow the transformation of complex parameter spaces into interpretable structures and enable direct comparison with features reported in the literature (Garcia-Dias et al. 2018, 2019). In this study, spectroscopic data from APOGEE DR17 and astrometric data from Gaia were combined to investigate stellar populations and merger signatures in the Milky Way. After selecting giant stars with reliable measurements and applying strict quality cuts, orbital parameters were computed under the McMillan potential (McMillan 2017) using the galpy package (Bovy 2015). To mitigate the dominance of disk stars in the sample, a selection cut of $e_p > 0.25$ was applied, following (Fernandes et al. 2023). The classification performance was then tested in three distinct parameter spaces: (i) chemical abundances only, (ii) dynamical parameters only, and (iii) a combination of both. The resulting clusters were evaluated in chemical abundance diagrams and Lindblad diagrams to assess their astrophysical interpretability. The results demonstrate that the clearest separation is achieved when using chemical parameters alone. This confirms the robustness of chemical signatures in tracing accreted populations, whereas purely dynamical or chemo-dynamical parameter sets tend to produce contamination and ambiguities in classification. Overall, the methodology presented here provides reproducible and data-driven criteria forstellar population classification. In future work, the approach will be extended by incorporating additional chemical abundances, which will further improve the accuracy of population separation and contribute valuable insights to the field of Galactic archeology.

2. Data and Methodology

We cross-matched APOGEE DR17 with Gaia DR3, applying quality cuts (S/N $\geq$ 70, $\sigma_{\varpi}/\varpi \leq 0.2$, and clean flags for ASPCAP, STAR, and Fe, Mg, Mn, Al, Ni, Si). Restricting to giants ($0.5 \leq \log g \leq 3.6$, $3500 \leq T_{\rm eff} \leq 5500$ K (Horta et al. 2023) yielded 135,080 stars. Using Astropy (Astropy Collaboration et al. 2013, 2018) and galpy (Bovy 2015), we derived stellar kinematics and orbits in the McMillan potential (McMillan 2017), adopting $R_{\odot} = 8.2$ kpc, $V_{\rm LSR} = 232.8$ km/s, and $(U, V, W) = (11.10, 12.24, 7.25)$ km/s (Schönrich et al. 2010), from which $R_{\rm apo}$, $R_{\rm peri}$, $Z_{\rm max}$, L_z, E, and e were obtained.

Since disk stars dominate the sample, and this may introduce bias into the machine learning procedure, we followed the approach of (Fernandes et al. 2023) and restricted the analysis to stars with $e_p > 0.25$. Subsequently, we employed PCA for dimensionality reduction using three different sets of input features-chemical abundances, kinematic parameters, and their combination. The dimensionality was reduced by monitoring inertia values, after which we applied the K-means clustering algorithm. While the number of clusters was initially guided by inertia breakpoints, inspection of the resulting distributions in the [Mg/Mn]-[Al/Fe] and E-L_z planes revealed that not all breakpoints yielded astrophysically meaningful results, a difficulty also noted by (Plevne and Akbaba 2025). Therefore, we retained only those cluster numbers that produced physically interpretable outcomes, and examined the corresponding stellar distributions in the relevant planes

https://doi.org/10.1017/S1743921326106255 Published online by Cambridge University Press

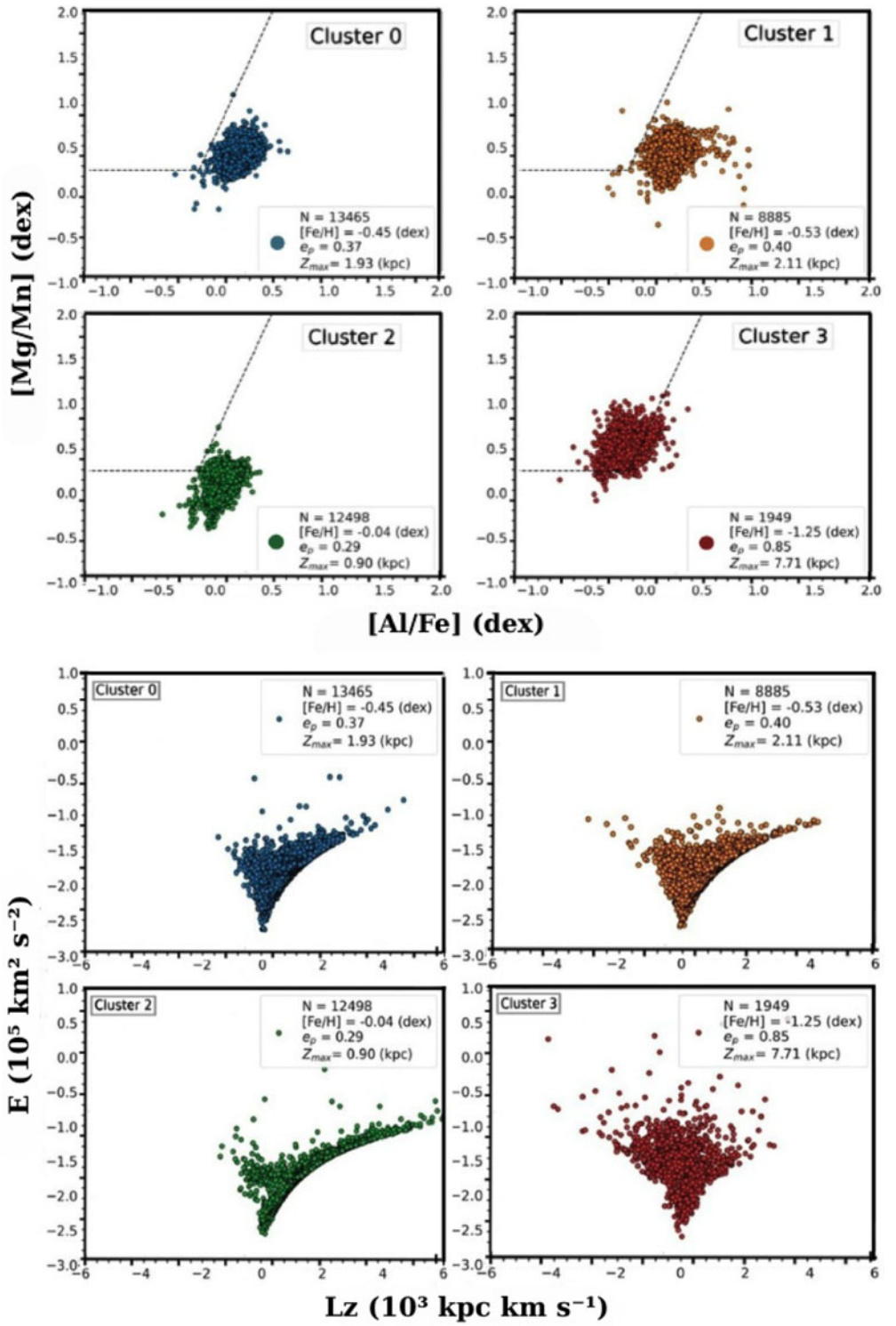


Figure 1. Clustering results of giant stars using chemical abundances as input features. The panels show the distributions in the [Mg/Mn]-[Al/Fe] and E-L_z planes. For each cluster, the stellar counts, along with median values of [Fe/H], eccentricity (e), and $Z_{\max}$, are indicated. Despite some overlap, the clusters correspond to the thin disk (Cluster 2), thick disk (Cluster 0), halo (Cluster 1), and accreted stars (Cluster 3).

https://doi.org/10.1017/S1743921326106255 Published online by Cambridge University Press

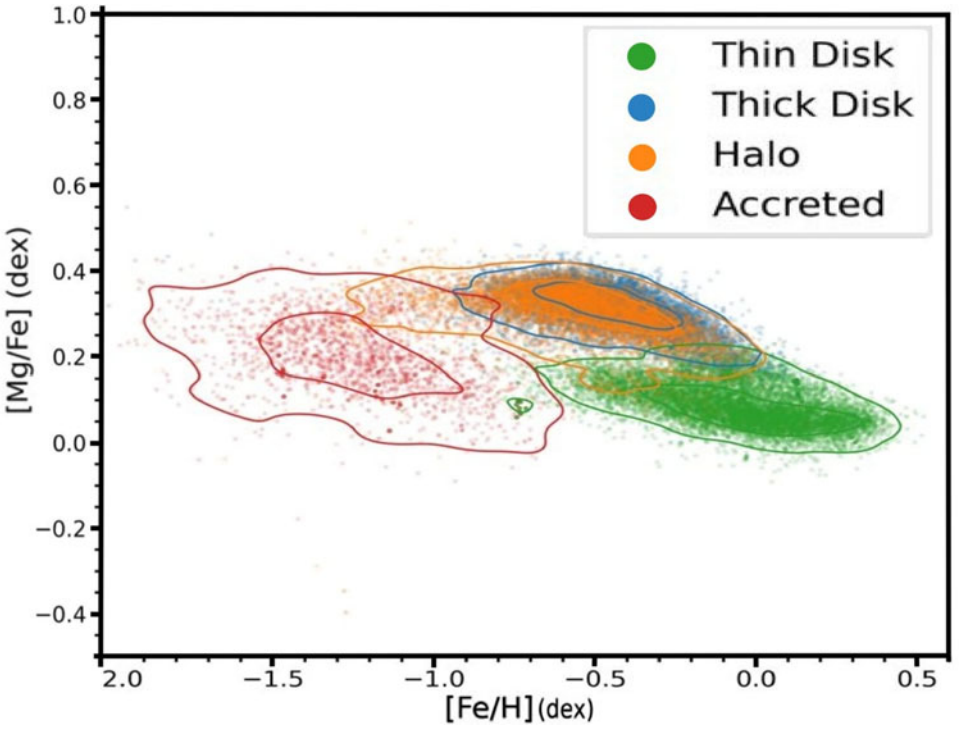


Figure 2. Distribution of the four clusters obtained with PCA + K-means in the [Mg/Fe]-[Fe/H] plane. The clusters reveal distinct loci corresponding to the thin disk, the thick disk/halo, and the accreted population, with partial overlap between components.

3. PCA and K-Means Clustering Algorithm Results

For the dimensionality reduction applied to giant stars, we examined the relationship between cumulative explained variance and the number of components to determine the dimensional space that yields meaningful results. Following this, we applied the K-means clustering algorithm with inertia monitoring. For the chemical-abundance input space, we used [Fe/H], [Mg/Fe], [Mn/Fe], [Al/Fe], [Ni/Fe], and [Si/Fe]; for the kinematic input space, we adopted $S_{\rm LSR}$, E, L_z, v_ϕ, $W_{\rm LSR}$, $Z_{\rm max}$, v_z, v_r, $R_{\rm apo}$, $R_{\rm peri}$, and e; and for the combined case, we included all of these parameters. The results indicate that the most robust approach is obtained when using only the chemical abundances as input features. As shown in Figure 1, the stellar counts in each cluster, together with their median metallicities, eccentricities, and $Z_{\rm max}$ values, are reported for both the [Mg/Mn]-[Al/Fe] and E-L_z planes. Although some overlap is present, the clusters can be interpreted as follows: Cluster 2 corresponds to the thin disk, Cluster 0 to the thick disk, Cluster 1 to the halo, and Cluster 3 to the accreted population. Despite the $e_p > 0.25$ cut, the dominance of disk stars remains visible.

4. Conclusion

By combining APOGEE DR17 and Gaia DR3 data, we constructed a high-quality giant- star sample and applied PCA with K-means to explore the chemo-dynamical structure of the Galaxy. Tests with different input spaces showed that chemical abundances alone provide the clearest separation of stellar populations. In particular, as illustrated in Figure 2, the [Mg/Fe]-[Fe/H] plane reveals distinct loci for the thin disk, the thick

https://doi.org/10.1017/S1743921326106255 Published online by Cambridge University Press

disk/halo, and the accreted component. Overall, our results demonstrate that abundance-driven clustering is an effective tool for disentangling Galactic populations and isolating ex-situ stars.

Acknowledgements

This research was supported by the Scientific and Technological Research Council of Turkey (TÜBİTAK) under project number 123F227 and the 2211-C Priority Areas Fellowship Program. Additional funding was provided by the Scientific Research Projects Coordination Unit of Istanbul University under project numbers 40044, FDK-2025-41537.

References

Abdurro'uf, Accetta, K., Aerts, C., Silva Aguirre, V., Ahumada, R., Ajgaonkar, N., Filiz Ak, N., Alam, S., Allende Prieto, C., Almeida, A., Anders, F., Anderson, S. F., Andrews, B. H., Anguiano, B., Aquino-Ortíz, E., Aragón-Salamanca, A., Argudo-Fernández, M., Ata, M., Aubert, M., Avila-Reese, V., Badenes, C., Barbá, R. H., Barger, K., Barrera-Ballesteros, J. K., Beaton, R. L., Beers, T. C., Belfiore, F., Bender, C. F., Bernardi, M., Bershady, M. A., Beutler, F., Bidin, C. M., Bird, J. C., Bizyaev, D., Blanc, G. A., Blanton, M. R., Boardman, N. F., Bolton, A. S., Boquien, M., Borissova, J., Bovy, J., Brandt, W. N., Brown, J., Brownstein, J. R., Brusa, M., Buchner, J., Bundy, K., Burchett, J. N., Bureau, M., Burgasser, A., Cabang, T. K., Campbell, S., Cappellari, M., Carlberg, J. K., Wanderley, F. C., Carrera, R., Cash, J., Chen, Y.-P., Chen, W.-H., Cherinka, B., Chiappini, C., Choi, P. D., Chojnowski, S. D., Chung, H., Clerc, N., Cohen, R. E., Comerford, J. M., Comparat, J., da Costa, L., Covey, K., Crane, J. D., Cruz-Gonzalez, I., Culhane, C., Cunha, K., Dai, Y. S., Damke, G., Darling, J., Davidson, James W., J., Davies, R., Dawson, K., De Lee, N., Diamond-Stanic, A. M., Cano-Díaz, M., Sánchez, H. D., Donor, J., Duckworth, C., Dwelly, T., Eisenstein, D. J., Elsworth, Y. P., Emsellem, E., Eracleous, M., Escoffier, S., Fan, X., Farr, E., Feng, S., Fernández-Trincado, J. G., Feuillet, D., Filipp, A., Fillingham, S. P., Frinchaboy, P. M., Fromenteau, S., Galbany, L., García, R. A., García-Hernández, D. A., Ge, J., Geisler, D., Gelfand, J., Géron, T., Gibson, B. J., Goddy, J., Godoy-Rivera, D., Grabowski, K., Green, P. J., Greener, M., Grier, C. J., Griffith, E., Guo, H., Guy, J., Hadjara, M., Harding, P., Hasselquist, S., Hayes, C. R., Hearty, F., Hernández, J., Hill, L., Hogg, D. W., Holtzman, J. A., Horta, D., Hsieh, B.-C., Hsu, C.-H., Hsu, Y.-H., Huber, D., Huertas-Company, M., Hutchinson, B., Hwang, H. S., Ibarra-Medel, H. J., Chitham, J. I., Ilha, G. S., Imig, J., Jaekle, W., Jayasinghe, T., Ji, X., Johnson, J. A., Jones, A., Jönsson, H., Katkov, I., Khalatyan, Arman, D., Kinemuchi, K., Kisku, S., Knapen, J. H., Kneib, J.-P., Kollmeier, J. A., Kong, M., Kounkel, M., Kreckel, K., Krishnarao, D., Lacerna, I., Lane, R. R., Langgin, R., Lavender, R., Law, D. R., Lazarz, D., Leung, H. W., Leung, H.-H., Lewis, H. M., Li, C., Li, R., Lian, J., Liang, F.-H., Lin, L., Lin, Y.-T., Lin, S., Lintott, C., Long, D., Longa-Peña, P., López-Cobá, C., Lu, S., Lundgren, B. F., Luo, Y., Mackereth, J. T., de la Macorra, A., Mahadevan, S., Majewski, S. R., Manchado, A., Mandeville, T., Maraston, C., Margalef-Bentabol, B., Masseron, T., Masters, K. L., Mathur, S., McDermid, R. M., Mckay, M., Merloni, A., Merrifield, M., Meszaros, S., Miglio, A., Di Mille, F., Minniti, D., Minsley, R., Monachesi, A., Moon, J., Mosser, B., Mulchaey, J., Muna, D., Muñoz, R. R., Myers, A. D., Myers, N., Nadathur, S., Nair, P., Nandra, K., Neumann, J., Newman, J. A., Nidever, D. L., Nikakhtar, F., Nitschelm, C., O'Connell, J. E., Garma-Oehmichen, L., Luan Souza de Oliveira, G., Olney, R., Oravetz, D., Ortigoza-Urdaneta, M., Osorio, Y., Otter, J., Pace, Z. J., Padilla, N., Pan, K., Pan, H.-A., Parikh, T., Parker, J., Peirani, S., Peña Ramírez, K., Penny, S., Percival, W. J., Perez-Fournon, I., Pinsonneault, M., Poidevin, F., Poovelil, V. J., Price-Whelan, A. M., Bárbara de Andrade Queiroz, A., Raddick, M. J., Ray, A., Rembold, S. B., Riddle, N., Riffel, R. A., Riffel, R., Rix, H.-W., Robin, A. C., Rodríguez-Puebla, A., Roman-Lopes, A., Román-Zúñiga, C., Rose, B., Ross, A. J., Rossi, G., Rubin, K. H. R., Salvato, M., Sánchez, S. F., Sánchez-Gallego, J. R., Sanderson, R., Santana Rojas, F. A., Sarceno, E., Sarmiento, R., Sayres, C., Sazonova, E.,

https://doi.org/10.1017/S1743921326106255 Published online by Cambridge University Press

Schaefer, A. L., Schiavon, R., Schlegel, D. J., Schneider, D. P., Schultheis, M., Schwope, A., Serenelli, A., Serna, J., Shao, Z., Shapiro, G., Sharma, A., Shen, Y., Shetrone, M., Shu, Y., Simon, J. D., Skrutskie, M. F., Smethurst, R., Smith, V., Sobeck, J., Spoo, T., Sprague, D., Stark, D. V., Stassun, K. G., Steinmetz, M., Stello, D., Stone-Martinez, A., Storchi-Bergmann, T., Stringfellow, G. S., Stutz, A., Su, Y.-C., Taghizadeh-Popp, M., Talbot, M. S., Tayar, J., Telles, E., Teske, J., Thakar, A., Theissen, C., 2022, The Seventeenth Data Release of the Sloan Digital Sky Surveys: Complete Release of MaNGA, MaStar, and APOGEE-2 Data. *ApJS*, 259(2), 35.

Astropy Collaboration, Price-Whelan, A. M., Sipőcz, B. M., Günther, H. M., Lim, P. L., Crawford, S. M., Conseil, S., Shupe, D. L., Craig, M. W., Dencheva, N., Ginsburg, A., VanderPlas, J. T., Bradley, L. D., Pérez-Suárez, D., de Val-Borro, M., Aldcroft, T. L., Cruz, K. L., Robitaille, T. P., Tollerud, E. J., Ardelean, C., Babej, T., Bach, Y. P., Bachetti, M., Bakanov, A. V., Bamford, S. P., Barentsen, G., Barmby, P., Baumbach, A., Berry, K. L., Biscani, F., Boquien, M., Bostroem, K. A., Bouma, L. G., Brammer, G. B., Bray, E. M., Breytenbach, H., Buddelmeijer, H., Burke, D. J., Calderone, G., Cano Rodríguez, J. L., Cara, M., Cardoso, J. V. M., Cheedella, S., Copin, Y., Corrales, L., Crichton, D., D'Avella, D., Deil, C., Depagne, É., Dietrich, J. P., Donath, A., Droettboom, M., Earl, N., Erben, T., Fabbro, S., Ferreira, L. A., Finethy, T., Fox, R. T., Garrison, L. H., Gibbons, S. L. J., Goldstein, D. A., Gommers, R., Greco, J. P., Greenfield, P., Groener, A. M., Grollier, F., Hagen, A., Hirst, P., Homeier, D., Horton, A. J., Hosseinzadeh, G., Hu, L., Hunkeler, J. S., Ivezić, Ž., Jain, A., Jenness, T., Kanarek, G., Kendrew, S., Kern, N. S., Kerzendorf, W. E., Khvalko, A., King, J., Kirkby, D., Kulkarni, A. M., Kumar, A., Lee, A., Lenz, D., Littlefair, S. P., Ma, Z., Macleod, D. M., Mastropietro, M., McCully, C., Montagnac, S., Morris, B. M., Mueller, M., Mumford, S. J., Muna, D., Murphy, N. A., Nelson, S., Nguyen, G. H., Ninan, J. P., Nöthe, M., Ogaz, S., Oh, S., Parejko, J. K., Parley, N., Pascual, S., Patil, R., Patil, A. A., Plunkett, A. L., Prochaska, J. X., Rastogi, T., Reddy Janga, V., Sabater, J., Sakurikar, P., Seifert, M., Sherbert, L. E., Sherwood-Taylor, H., Shih, A. Y., Sick, J., Silbiger, M. T., Singanamalla, S., Singer, L. P., Sladen, P. H., Sooley, K. A., Sornarajah, S., Streicher, O., Teuben, P., Thomas, S. W., Tremblay, G. R., Turner, J. E. H., Terrón, V., van Kerkwijk, M. H., de la Vega, A., Watkins, L. L., Weaver, B. A., Whitmore, J. B., Woillez, J., Zabalza, V., & Astropy Contributors 2018, The Astropy Project: Building an Open-science Project and Status of the v2.0 Core Package. *AJ*, 156(3), 123.

Astropy Collaboration, Robitaille, T. P., Tollerud, E. J., Greenfield, P., Droettboom, M., Bray, E., Aldcroft, T., Davis, M., Ginsburg, A., Price-Whelan, A. M., Kerzendorf, W. E., Conley, A., Crighton, N., Barbary, K., Muna, D., Ferguson, H., Grollier, F., Parikh, M. M., Nair, P. H., Unther, H. M., Deil, C., Woillez, J., Conseil, S., Kramer, R., Turner, J. E. H., Singer, L., Fox, R., Weaver, B. A., Zabalza, V., Edwards, Z. I., Azalee Bostroem, K., Burke, D. J., Casey, A. R., Crawford, S. M., Dencheva, N., Ely, J., Jenness, T., Labrie, K., Lim, P. L., Pierfederici, F., Pontzen, A., Ptak, A., Refsdal, B., Servillat, M., & Streicher, O. 2013, Astropy: A community Python package for astronomy. *A&A*, 558, A33.

Barbá, R. H., Minniti, D., Geisler, D., Alonso-García, J., Hempel, M., Monachesi, A., Arias, J. I., & Gómez, F. A. 2019, A Sequoia in the Garden: FSR 1758—Dwarf Galaxy or Giant Globular Cluster? *ApJ*, 870(2), L24.

Belokurov, V., Erkal, D., Evans, N. W., Koposov, S. E., & Deason, A. J. 2018, Co-formation of the disc and the stellar halo. *MNRAS*, 478(1), 611–619.

Belokurov, V. & Kravtsov, A. 2024, In-situ versus accreted Milky Way globular clusters: a new classification method and implications for cluster formation. *MNRAS*, 528(2), 3198–3216.

Bland-Hawthorn, J. & Gerhard, O. 2016, The Galaxy in Context: Structural, Kinematic, and Integrated Properties. *ARA&A*, 54, 529–596.

Bovy, J. 2015, galpy: A python Library for Galactic Dynamics. *ApJS*, 216(2), 29.

Şentürk, S. A., Şahin, T., Güney, F., Bilir, S., & Marışmak, M. 2024, Near-infrared Spectroscopy of the Sun and Solar Analog Star HD 76151: Compiling an Extensive Line List in the Y, J, H, and K Bands. *ApJ*, 976(2), 175.

Fernandes, L., Mason, A. C., Horta, D., Schiavon, R. P., Hayes, C., Hasselquist, S., Feuillet, D., Beaton, R. L., Jönsson, H., Kisku, S., Lacerna, I., Lian, J., Minniti, D., & Villanova, S.

https://doi.org/10.1017/S1743921326106255 Published online by Cambridge University Press

2023, A comparative analysis of the chemical compositions of Gaia-Enceladus/Sausage and Milky Way satellites using APOGEE. *MNRAS*, 519(3), 3611–3622.

Gaia Collaboration, Vallenari, A., Brown, A. G. A., Prusti, T., de Bruijne, J. H. J., Arenou, F., Babusiaux, C., Biermann, M., Creevey, O. L., Ducourant, C., Evans, D. W., Eyer, L., Guerra, R., Hutton, A., Jordi, C., Klioner, S. A., Lammers, U. L., Lindegren, L., Luri, X., Mignard, F., Panem, C., Pourbaix, D., Randich, S., Sartoretti, P., Soubiran, C., Tanga, P., Walton, N. A., Bailer-Jones, C. A. L., Bastian, U., Drimmel, R., Jansen, F., Katz, D., Lattanzi, M. G., van Leeuwen, F., Bakker, J., Cacciari, C., Castañeda, J., De Angeli, F., Fabricius, C., Fouesneau, M., Frémat, Y., Galluccio, L., Guerrier, A., Heiter, U., Masana, E., Messineo, R., Mowlavi, N., Nicolas, C., Nienartowicz, K., Pailler, F., Panuzzo, P., Riclet, F., Roux, W., Seabroke, G. M., Sordo, R., Thévenin, F., Gracia-Abril, G., Portell, J., Teyssier, D., Altmann, M., Andrae, R., Audard, M., Bellas-Velidis, I., Benson, K., Berthier, J., Blomme, R., Burgess, P. W., Busonero, D., Busso, G., Cánovas, H., Carry, B., Cellino, A., Cheek, N., Clementini, G., Damerdji, Y., Davidson, M., de Teodoro, P., Nuñez Campos, M., Delchambre, L., Dell'Oro, A., Esquej, P., Fernández-Hernández, J., Fraile, E., Garabato, D., García-Lario, P., Gosset, E., Haigron, R., Halbwachs, J. L., Hambly, N. C., Harrison, D. L., Hernández, J., Hestroffer, D., Hodgkin, S. T., Holl, B., Janßen, K., Jevardat de Fombelle, G., Jordan, S., Krone-Martins, A., Lanzafame, A. C., Löffler, W., Marchal, O., Marrese, P. M., Moitinho, A., Muinonen, K., Osborne, P., Pancino, E., Pauwels, T., Recio-Blanco, A., Reylé, C., Riello, M., Rimoldini, L., Roegiers, T., Rybizki, J., Sarro, L. M., Siopis, C., Smith, M., Sozzetti, A., Utrilla, E., van Leeuwen, M., Abbas, U., Ábrahám, P., Abreu Aramburu, A., Aerts, C., Aguado, J. J., Ajaj, M., Aldea-Montero, F., Altavilla, G., Álvarez, M. A., Alves, J., Anders, F., Anderson, R. I., Anglada Varela, E., Antoja, T., Baines, D., Baker, S. G., Balaguer-Núñez, L., Balbinot, E., Balog, Z., Barache, C., Barbato, D., Barros, M., Barstow, M. A., Bartolomé, S., Bassilana, J. L., Bauchet, N., Becciani, U., Bellazzini, M., Berihuete, A., Bernet, M., Bertone, S., Bianchi, L., Binnenfeld, A., Blanco-Cuaresma, S., Blazere, A., Boch, T., Bombrun, A., Bossini, D., Bouquillon, S., Bragaglia, A., Bramante, L., Breedt, E., Bressan, A., Brouillet, N., Brugaletta, E., Bucciarelli, B., Burlacu, A., Butkevich, A. G., Buzzi, R., Caffau, E., Cancelliere, R., Cantat-Gaudin, T., Carballo, R., Carlucci, T., Carnerero, M. I., Carrasco, J. M., Casamiquela, L., Castellani, M., Castro-Ginard, A., Chaoul, L., Charlot, P., Chemin, L., Chiaramida, V., Chiavassa, A., Chornay, N., Comoretto, G., Contursi, G., Cooper, W. J., Cornez, T., Cowell, S., Crifo, F., Cropper, M., Crosta, M., Crowley, C., Dafonte, C., Dapergolas, A., David, M., David, P., de Laverny, P., De Luise, F., De March, R., De Ridder, J., de Souza, R., de Torres, A., del Peloso, E. F., del Pozo, E., Delbo, M., Delgado, A., Delisle, J. B., Demouchy, C., Dharmawardena, T. E., Di Matteo, P., Diakite, S., Diener, C., Distefano, E., Dolding, C., Edvardsson, B., Enke, H., Fabre, C., Fabrizio, M., Faigler, S., Fedorets, G., Fernique, P., Fienga, A., Figueras, F., Fournier, Y., Fouron, C., Fragkoudi, F., Gai, M., Garcia-Gutierrez, A., Garcia-Reinaldos, M., García-Torres, M., Garofalo, A., Gavel, A., Gavras, P., Gerlach, E., Geyer, R., Giacobbe, P., Gilmore, G., Girona, S., Giuffrida, G., Gomel, R., Gomez, A., González-Núñez, J., González-Santamaría, I., González-Vidal, J. J., Granvik, M., Guillout, P., Guiraud, J., Gutiérrez-Sánchez, R., Guy, L. P., Hatzidimitriou, D., Hauser, M., Haywood, M., Helmer, A., Helmi, A., Sarmiento, M. H., Hidalgo, S. L., Hilger, T., Hładczuk, N., Hobbs, D., Holland, G., Huckle, H. E., Jardine, K., Jasniewicz, G., Jean-Antoine Piccolo, A., Jiménez-Arranz, Ó., Jorissen, A., Juaristi Campillo, J., Julbe, F., Karbevska, L., Kervella, P., Khanna, S., Kontizas, M., Kordopatis, G., Korn, A. J., Kóspál, Á., Kostrzewa-Rutkowska, Z., Kruszyńska, K., Kun, M., Laizeau, P., Lambert, S., Lanza, A. F., Lasne, Y., Le Campion, J. F., Lebreton, Y., Lebzelter, T., Leccia, S., Leclerc, N., Lecoeur-Taibi, I., Liao, S., Licata, E. L., Lindstrøm, H. E. P., Lister, T. A., Livanou, E., Lobel, A., Lorca, A., Loup 2023, Gaia Data Release 3. Summary of the content and survey properties. *A&A*, 674, A1.

Garcia-Dias, R., Allende Prieto, C., Sánchez Almeida, J., & Alonso Palicio, P. 2019, Machine learning in APOGEE. Identification of stellar populations through chemical abundances. *A&A*, 629, A34.

https://doi.org/10.1017/S1743921326106255 Published online by Cambridge University Press

Garcia-Dias, R., Allende Prieto, C., Sánchez Almeida, J., & Ordovás-Pascual, I. 2018, Machine learning in APOGEE. Unsupervised spectral classification with K-means. *A&A*, 612, A98.

Haywood, M., Di Matteo, P., Lehnert, M. D., Snaith, O., Khoperskov, S., & Gómez, A. 2018, In Disguise or Out of Reach: First Clues about In Situ and Accreted Stars in the Stellar Halo of the Milky Way from Gaia DR2. *ApJ*, 863(2), 113.

Helmi, A., Babusiaux, C., Koppelman, H. H., Massari, D., Veljanoski, J., & Brown, A. G. A. 2018, The merger that led to the formation of the Milky Way's inner stellar halo and thick disk. *Nature*, 563(7729), 85–88.

Helmi, A., White, S. D. M., de Zeeuw, P. T., & Zhao, H. 1999, Debris streams in the solar neighbourhood as relicts from the formation of the Milky Way. *Nature*, 402(6757), 53–55.

Horta, D., Schiavon, R. P., Mackereth, J. T., Weinberg, D. H., Hasselquist, S., Feuillet, D., O'Connell, R. W., Anguiano, B., Allende-Prieto, C., Beaton, R. L., Bizyaev, D., Cunha, K., Geisler, D., García-Hernández, D. A., Holtzman, J., Jönsson, H., Lane, R. R., Majewski, S. R., Mészáros, S., Minniti, D., Nitschelm, C., Shetrone, M., Smith, V. V., & Zasowski, G. 2023, The chemical characterization of halo substructure in the Milky Way based on APOGEE. *MNRAS*, 520(4), 5671–5711.

Ibata, R. A., Gilmore, G., & Irwin, M. J. 1994, A dwarf satellite galaxy in Sagittarius. *Nature*, 370(6486), 194–196.

Jin, X. & Han, J. 2011. *K-Means Clustering*, pp. 563–564.

Jolliffe, I. 1990, Principal component analysis: A beginner's guide - i. introduction and application. *Weather*, 45, 375–382.

Koppelman, H. H., Helmi, A., Massari, D., Price-Whelan, A. M., & Starkenburg, T. K. 2019, Multiple retrograde substructures in the Galactic halo: A shattered view of Galactic history. *A&A*, 631, L9.

Mackereth, J. T., Schiavon, R. P., Pfeffer, J., Hayes, C. R., Bovy, J., Anguiano, B., Allende Prieto, C., Hasselquist, S., Holtzman, J., Johnson, J. A., Majewski, S. R., O'Connell, R., Shetrone, M., Tissera, P. B., & Fernández-Trincado, J. G. 2019, The origin of accreted stellar halo populations in the Milky Way using APOGEE, Gaia, and the EAGLE simulations. *MNRAS*, 482(3), 3426–3442.

Malhan, K., Ibata, R. A., Sharma, S., Famaey, B., Bellazzini, M., Carlberg, R. G., D'Souza, R., Yuan, Z., Martin, N. F., & Thomas, G. F. 2022, The Global Dynamical Atlas of the Milky Way Mergers: Constraints from Gaia EDR3-based Orbits of Globular Clusters, Stellar Streams, and Satellite Galaxies. *ApJ*, 926(2), 107.

Matsuno, T., Aoki, W., & Suda, T. 2019, Origin of the Excess of High-energy Retrograde Stars in the Galactic Halo. *ApJ*, 874(2), L35.

McMillan, P. J. 2017, The mass distribution and gravitational potential of the Milky Way. *MNRAS*, 465(1), 76–94.

Myeong, G. C., Vasiliev, E., Iorio, G., Evans, N. W., & Belokurov, V. 2019, Evidence for two early accretion events that built the Milky Way stellar halo. *MNRAS*, 488(1), 1235–1247.

Naidu, R. P., Conroy, C., Bonaca, A., Johnson, B. D., Ting, Y.-S., Caldwell, N., Zaritsky, D., & Cargile, P. A. 2020, Evidence from the H3 Survey That the Stellar Halo Is Entirely Comprised of Substructure. *ApJ*, 901(1), 48.

Plevne, O. & Akbaba, F. 2025, Estimating the Initial Mass of the Gaia-Enceladus Dwarf Galaxy with a Chemical Evolution Model. *ApJ*, 991(2), 207.

Plevne, O., Önal Tas, Ö., Bilir, S., & Seabroke, G. M. 2020, Multiwavelength Absolute Magnitudes and Colors of Red Clump Stars in the Gaia Era. *ApJ*, 893(2), 108.

Schönrich, R., Binney, J., & Dehnen, W. 2010, Local kinematics and the local standard of rest. *MNRAS*, 403(4), 1829–1833.

Yuan, Z., Chang, J., Beers, T. C., & Huang, Y. 2020, A Low-mass Stellar-debris Stream Associated with a Globular Cluster Pair in the Halo. *ApJ*, 898(2), L37.

https://doi.org/10.1017/S1743921326106255 Published online by Cambridge University Press